

FODOMUST - Une plateforme de clustering collaboratif sous contraintes incrémental de séries temporelles

Pierre Gañçarski, Baptiste Lafabregue, Abdoul-djawadou Salaou, Harrison Vernier

ICube, CNRS - Université de Strasbourg
{gancarski, lafabregue, adsalaou, harrison.vernier}@unistra.fr

Résumé. La plateforme FODOMUST¹ est une implantation concrète des méthodes, bibliothèques et interfaces dédiées au clustering de données complexes proposées au sein d'ICube. Elle intègre une version multisource de la méthode de clustering collaboratif multistratégie SAMARAH. Sa principale originalité est qu'elle offre à l'utilisateur les méthodes et les interfaces permettant le clustering sous contraintes incrémental de données temporelles symboliques ou numériques. L'interface MULTICUBE dédiée à l'analyse de séries temporelles offre aussi un accès à un ensemble d'algorithmes de segmentation soit propres à ICube soit faisant appel à l'OTB diffusée par le CNES.

1 Introduction

Les méthodes d'apprentissage supervisé classiquement utilisées font l'hypothèse que les données d'apprentissage décrivent de manière suffisante et complète les classes auxquelles elles sont rattachées. En d'autres termes, ces méthodes nécessitent que les classes recherchées soient parfaitement connues et définies et que l'expert soit capable de fournir un jeu de données d'apprentissage suffisant tant en nombre qu'en qualité. Or, face aux flux quasi-continus de nouvelles données issues de capteurs de plus en plus nombreux et variés, cette hypothèse n'est plus réaliste dans bien des domaines. En effet, la révolution technologique de la haute fréquence d'acquisition y est encore trop récente pour que les connaissances thématiques se soient adaptées. Ainsi, bien souvent il n'existe pas de typologies (ou nomenclatures) des changements réellement utilisables pour ce type d'analyse supervisée et donc de données d'apprentissage de qualité associées. Pour pallier ce manque, des approches récentes basées sur un apprentissage non supervisé font l'hypothèse que même en l'absence de connaissances formalisées, celles-ci peuvent néanmoins être représentées en partie à travers des contraintes (de comparaison, d'étiquetage ou de structure) opérables (Basu et al., 2008; Dao et al., 2017; Lampert et al., 2018). Ces contraintes a priori plus aisées à définir, peuvent alors être utilisées pour guider le processus de clustering afin de produire des clusters plus proches des « intuitions » de l'expert c'est-à-dire des classes thématiques potentielles. Ainsi, la méthode collaborative SAMARAH (Gañçarski et Wemmert, 2007; Forestier et al., 2010) développée par ICube a été complétée par une prise en compte incrémentale de contraintes (Lampert et al., 2019).

1. <http://icube-sdc.unistra.fr/fr/index.php/Plateformes>